DISCUSSION AND RE-ANALYSIS OF EXPERIMENTAL DATA IN THE INVESTIGATION OF ULTRADIAN RHYTHMS IN HUMANS

Torsten Andresen

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

DISCUSSION AND RE-ANALYSIS OF EXPERIMENTAL DATA IN THE INVESTIGATION OF ULTRADIAN RHYTHMS IN HUMANS

by

Torsten Andresen

March 1981

Thesis Advisors: Douglas E. Neil
P. A. W. Lewis

Approved for public release; distribution unlimited

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Discussion and Re-Analysis of Experimental Data in the Investigation of Ultradian Rhythms in Humans		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis March, 1981
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Torsten Andresen		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March, 1981
		13. NUMBER OF PAGES 62
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ultradian Rhythm

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper a discussion and re-analysis of experimental data is carried out. The data consist of measurements of the performance on two tasks (verbal and spatial) that are performed by the two different brain halves. The original experiments and analysis had been conducted by R. Klein and R. Armitage in the investigation of ultradian rhythms in humans. The period of the ultradian cycle had been hypothetically equated to the period of REM-sleep occurrence. In the re-analysis, previously unmentioned inhomogeneity was discovered and the findings by the original experimenter-analysts could not be

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (cont'd) confirmed at the stated significance levels. A discussion of and recommendations for the experimental set-up are included. The re-analysis was conducted with non-parametric methods. Small ultradian rhythmicity was concluded to have been induced by influence of circadian rhythm, experimental set-up and inhomogeneity among subjects.

Discussion and Re-analysis of Experimental Data in the Investigation of Ultradian Rhythms in Humans

by

Torsten Andresen
Kapitaenleutnant, Federal German Navy

Submitted in partial fullfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1981

ABSTRACT

In this paper a discussion and re-analysis of experimental data is carried out. The data consist of measurements of the performance on two tasks (verbal and spatial) that are performed by the two different brain halves. The original experiments and analysis had been conducted by R. Klein and R. Armitage in the investigation of ultradian rhythms in humans. The period of the ultradian cycle had been hypothetically equated to the period of REM-sleep occurrence. In the re-analysis, previously unmentioned inhomogeneity was discovered and the findings by the original experimenter-analysts could not be confirmed at the stated significance levels. A discussion of and recommendations for the experimental set-up are included. The re-analysis was conducted with non-parametric methods. Small ultradian rhythmicity was concluded to have been induced by influence of circadian rhythm, experimental set-up and inhomogeneity among subjects.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. DISCUSSION

A. INTRODUCTION

An awareness and interest in various cyclic phenomena is not a recent development. Man, since his earliest existence, has been aware of cyclical influence in his environment: day changed to night, only to repeat itself about 12 hours later; spring, summer, fall and winter followed each other continuously; the recurrent changing of the moon, ebb and flow of tides and the female menstrual cycle were obvious. The Greeks, however, were perhaps the first to actually apply a knowledge of cycles in an attempt to better understand their environment. Luce [1970] suggests that the Greeks applied their knowledge of cycles in the treatment of certain illnesses over 2000 years ago.

This knowledge and interest was obscured during the period known as the Dark Ages. At the end of the 19th century a theory of biological cycles known as biorhythm rekindled the interest and stimulated research in the area. During the 20th century the interest has intensified with more cycles being identified and serious research activity devoted in an attempt to understand the nature of the various biological cycles.

B. DEFINITION AND EXPLANATION OF CYCLES

A cycle is a course or series of events that recur regularly and usually lead back to the starting point. [Webster, 1976] Mathematically, a function is cyclic with period T if $X(t+kT)=x(t)$ for $k=0,\pm 1,\pm 2,\ldots$

A rhythm is a regularly recurrant quantitative change in a biological process. [Webster, 1976]

A rhythm - as defined above - is produced by oscillators; these can be either linear or non-linear oscillators, thus causing linear (e.g. sinusoidal) or non-linear (e.g. relaxation, "saw tooth") oscillations.

Just as the cycles observed in man's environment since ancient times occur in different frequencies, the periods (1 over frequency) of rhythmic activities in biological systems vary from seconds or minutes (e.g. in cell growth and division) to hours (e.g. the 24-hour cycle of the circadian rhythm) to days or weeks (e.g. the 28-day clcle of female menstruation).

The terminology of cycle research is grouped around the most obvious one, the cycle with a period of "about a day" (about 24 hours). Dr Franz Halberg coined the term "circadian rhythm" for this from Latin "circa - about" and "dies - a day." Cycles with periods shorter than a day are called ultradian, those with periods longer than a day are refereed to as infradian cycles.

There are essentially two different possible explanations of rhythmic activities in biological systems. The first is that it is an essential dynamical feature of the process observed, i.e. it is part of the process and occurs within it at a certain place regardless of when and where the whole process takes place. The second is that such rhythms represent adaptive responses of the organism to a periodic environment; i.e. if the organism were not or had not been exposed to the periodicity (i.e. entrained by zeitgeber, a periodic clue-giving environmental feature), it would not show the rhythm. [Oatley and Goodwin, 1971]

An example for the first possible explanation - a dynamical feature regardless of the environment - is found in cell growth and division. If the cell division occured when triggered by the environment, it could easily happen that the resulting daughter cells would not receive a full set of the hereditary material (DNA), would then be deficient in genes and would not be able to survive. Considering the enormous number of cell divisions it takes to produce a new life, the probability of a healthy being to result from a cell division process that was environmentally induced is very small.

The second possible explanation - entrainment by zeitgeber - appears to account for all those cases that seem to have originated as a result of adaptation to a periodic environment. The most often discussed and rather obvious zeitgeber is the

24-hour day, held responsible for entrainment of the circadian rhythm; another is the 12-hour tidal cycle. Its entrainment has been studied in marine (littoral) organisms.

A most difficult to explain phenomenon occurs when the frequency of one internally caused cycle falls into the range of an environmentally induced cycle. The resulting complicated periodic organization then has to be taken as partly adaptive and partly of internal orgin. [Oatley and Goodwin, 1971] Bruce [1965] demonstrated that this is often the case for cell division of multicellular organisms. Generation times were found to converge in the neighborhood of 24 hours or a multiple thereof. Interaction of this kind may also occur when there are simple multiple relations between frequencies like the 24-hour cycle and the 28-day cycle.

C. CIRCADIAN AND ULTRADIAN RHYTHM RESEARCH DURING SLEEP

While psychological and physiological phenomena of the circadian as well as of the infradian rhythm have been researched extensively [Kleitmann, 1949; Harker, 1958; Aschoff, 1963; Lobban, 1965; Colquhoun, 1971], ultradian rhythms have only been studied during sleep-phases or in relation to sleep. Ultradian rhythms are those with a period shorter than 24 hours, a frequency thereby of more than one per day. With the use of electroencephalograph (EEG) that records voltages of the biological system through sensors connected to various areas of the skull, the depth of sleep -

among other things - can be measured. Not only have different levels of sleep been identified but it has also been found that one of the levels recurs in a cycle of about 90 minutes. This level or stage is the rapid eye movement stage (REM-stage), named after its physiological phenomenon. Another name for it is dream-sleep-stage, since it is during this stage that a person dreams. The REM cycle is the best researched of all ultradian cycles in animal and human observation. Hartmann [1967] quotes the average lengths in minutes as follows: mouse (3-4), rat (7-13), rabbit (24), opossum (17), cat (20-40), monkey (40-60), man (80-90), elephant(120). Kleitmann [1967] has attributed it to the change in metabolic rate. Be this as it may, the REM cycle during sleep is an excellent example of an ultradian rhythm.

D. ULTRADIAN RHYTHM RESEARCH DURING WAKEFULNESS

These findings accepted, the logical question to ask was whether a continuation of the REM cycle during the wakeful state could be hypothesized and observed. Kleitmann [1963] suggested that the basic rest-activity cycle of 80-90 minutes, as seen in the recurrant stages associated with dreaming in actual sleep, may persist during the waking period also, manifesting itself in recurrant fluctuations of alertness. [Colquhoun, 1971]

From there on it was only a question of time until scientists would look for parallels between the - by now well

researched - field of sleep and that of wakefulness. In 1922 a group of subjects had been kept in bed all day and a roughly 90-minute rhythm of body movement and stomach contraction had been observed. [Luce, 1970] Oral activity (eating, drinking, smoking) was found to increase and decrease with a range of 85-110 minutes with a mean of 96 minutes. [Friedman, 1965]

E. ASSOCIATED RESEARCH PHENOMENA

Scientific biological experimentation has always focused on animals. Hundreds of studies have been conducted observing anything from fruit flies to mice to monkeys. While the relative availability, short life cycles and possibility of complete destruction made observation and evaluation in animals easier than in human experiments, it also seems to have led astray many scientists and their audiences into the belief that primitive creatures certainly followed certain instincts and reflexes in leading their somewhat artificial lives. Many humans, however, would not readily accept that man himself, considered the crown of creation, still followed old innate or environmentally entrained oscillators that contribute to many of his emotions, actions, reactions and physiological phenomena.

It is also true that biological research in the field of cycles and rhythms is most difficult to perform on humans. As man subjects himself to externally imposed schedules of work, sleep, entertainment and food intake, the biologically "normal" cycle is often disturbed or willfully counteracted (e.g. by night work).

F. IMPACT AND IMPORTANCE OF RESEARCH

The importance of research in the field of rhythmic changes in human performance becomes obvious once one realizes how many actions, reactions and their effects or evaluations depend on human reaction time, accuracy, acuity, steadiness, etc., all potentially subject to variation via internal "biological clocks." [Kleitmann, 1949; Harker, 1958; Aschoff, 1963] While the circadian rhythm is generally accepted and - sometimes - considered in the scheduling of human activities (e.g. night shifts, watch bills), an ultradian rhythm has so far not been shown to occur during man's wakefulness and has therefore not been considered.

If it exists and its existence makes a significant difference in many or all of the above mentioned human activities and if it is predictable, the impact of its discovery could be extensive. Long term employment and work output may not be affected since in the long run (e.g. a day) the ups and downs would smoothe out to an average performance level that would follow the circadian rhythm. Any short term employment and output, however, could be affected significantly. School exams of one hour duration could fall either in the improved performance period or the degraded performance period of such a cycle; reaction to warning signs or signals could be impaired if the signal presentation occurred during the "down period."

The most significant immediate impact would probably be felt in the scientific community where tests of short duration

are very common. Medicine and medical research might have to revise some of their findings. For example, human reaction to drugs might be time-dependent or the effect of drugs could be over-shadowed by the effect of the ultradian rhythm. A drug that had been thought of as improving reaction time might be ineffective and any faster reaction time might have been due to an upswing in the ultradian rhythm. The effect of a drug that had been found ineffective might not have demonstrated desired results because a subject was on a downswing in the ultradian rhythm. As for the above mentioned drug experiments many subjects are used, it is not very likely that the discovery or acceptance of an ultradian rhythm would overthrow all findings, but it should be considered in the planning of future experiments, especially those that involve only few subjects and are of short duration.

An important question to ask before the above contemplations are considered, is whether performance is related to the ultradian rhythm - should it exist. Hunger (stomach contractions, sucking movements of babies), sex (penile erection during sleep or very tired wakefulness) and survival (REM sleep "alertness") are all neurologically based in the oldest part of the brain. Does performance ever draw from those parts or should it be influenced by them?

G. KLEIN'S AND ARMITAGE'S [1976] EXPERIMENTAL SET-UP

Based on the previously mentioned suggestions that a 90-minute cycle observed during sleep might continue in some form during the wake state, Klein and Armitage [1976] of the Department of Psychology at the Dalhousie University, Halifax, Canada, designed an experiment to investigate the hypothesis that the above suggestion was true. In addition to this hypothesis they also investigated previous findings by other scientists [Cohen, Gross; Geffen, et al. 1972] which indicated that the two brain halves - the right and left hemisphere - process different kinds of information. According to those findings the right hemisphere specialized in processing visual-spatial, wholistic, non-logical information, while the left half specialized in verbal-linquistic, analytic and logical processes. Combining these two areas of study the hypothesis under test was that performance measurements on each of two tasks - to be described below - should reveal rhythmic oscillations with a period of approximately 80-120 minutes and that equivalent phases of the two rhythms should be separated by 180° (one half cycle).

The experimenters chose two simultaneous matching tasks. For the spatial task subjects were required to match dot patterns of seven dots each, the verbal task required them to match letters. Two dot patterns had to be judged same or different and a capital letter was compared to a lower case letter and judged same or different. The actual recording of

the decision was made by the subject by crossing out S's or D's for same or different. The subjects were eight young adult volunteers in the age range of 18 to 24; there were five females and three males who were right-handed and had no left-handed siblings or parents.

All eight subjects were tested in a group testing room, all on the same day. They had two booklets in front of them on a desk, one for each task. The booklets contained 48 pages with 96 problems to a page. (The booklets were to be used over again after completion. An influence of the outcome by using the same problems over again was not to be expected since over 4000 problems would have to be solved before they repeated themselves).

The subjects were told to record their same/different choices as quickly as possible without making errors. They were given practice on each task to account for a possible learning curve (LC).

Performance was to be measured on each task every 15 minutes for three minutes. The test started at 0900 and ran for 8 hours, yielding 32 data points (observations) per task per subject. Lunch was from 1230 to 1300. The testing was not interrupted, but subjects ate in the free minutes during this period. Subjects were not allowed to talk to one another but could walk around, draw or read. They also reported in writing their thoughts at the end of every 15-minute interval. Dependent

variable for the test was number correctly matched pairs in each 3-minute test period. This was accepted to reflect the subjects' rate of accurate performance on the matching tasks, since the error rate was low (1.6%).

H. CONSIDERATIONS ABOUT AN ADVISED EXPERIMENTAL SET-UP

Before listing some general thoughts on experimental set-up for the kind of experiment as the one described in this thesis, it should be emphasized that by mentioning a consideration or recommending a procedure it is in no way implied that Klein and Armitage did not consider such question in the conduct of their effort. A few of the facts from their description of the experiment are contrary to recommendations that seem to be justified from the study of related research phenomena. Since it is impossible to describe an experimental set-up within a journal article of reasonable length and mention all reasoning behind each element of the set-up, any of the following recommendations about the conduct of such an experiment that are not in direct contrast to the way it actually was done should be understood more as an enlargement of the background knowledge than as an insinuation that Klein and Armitage might have neglected to consider them.

I. DISCUSSION OF EXPERIMENTAL SET-UP

To set up an experiment with expected outcomes of any significance several principles need to be observed:

1. The number of test subjects needs to be large enough to rule out findings by chance or idiosyncracies of the subjects chosen. For tests involving humans as subjects numbers have been as small as two in many experiments. This is very close to the borderline of unacceptability. For an experiment for which it is anticipated that the findings might be generalized to all humans, a representative mix of subjects has to be selected. Men, women, young, old,well educated and undereducated are some immediate distinctions. The last two categorizations might be especially critical when the experiment is a test and the variable tested is performance.
2. The test subjects are not to be informed of expected outcomes, but need to know in general how the test/ experiment is to be conducted. Without knowledge of procedures and measurements, human curiosity, anxiety or even fear could seriously contaminate test results.
3. The interaction between personnel administering the test elements and the Ss needs to be formalized and strictly standardized without the least chance for the Ss to "read between the lines" of any comments. Otherwise, the Ss will tend to "deliver" the desired result to the experimenter. [Rosenthal-effect; Russell, 1977]

4. Humans tend to act differently when they feel they are being ovserved than when they think they are unobserved. The increase in activity or productivity of working people due only to the fact that they are being "paid attention to" is known as the Hawthorne-effect. In performance-observation experiments this effect must not be neglected.
5. Constant presence and neutral behavior of the administering personnel are of paramount importance, if the experiment measures possible performance variation. If interspersed absence of the administering personnel is necessary, the times and lengths of absence need to be randomized and noted in relation to experimental outcomes.
6. Any regularity during an experiment directed to investigate rhythmic phenomena has to be avoided in order not to induce an artifical rhythm or cycle or a harmonic of it.
7. Physical separation of subjects would be helpful to ensure absolute non-interference between them. In tests like the one studied, where booklets have to be moved and pages are turned or moved, the simple sould of one S turning a page may be enough to evoke a "competitive spirit" in another. [Blake, 1970]

8. Immediately connected to avoidance of regularity and to separation of Ss, the next aspect of good experimental set-up would be temporal separation of the subjects. Not only would possible interaction between them be eliminated, but also the circadian-rhythm-and external everyday-life-cycle-induced regularities would be kept from contaminating the test results.
9. A different approach to the randomization of the times of test would be to use a smaller number of Ss - thereby increasing the homogeneity of the Ss - and have them perform on very similar tasks elements over several 8-hour-time blocks - beginning at different times - for several days to obtain at least the same amount of data points.
10. The biggest interference with any test of a length of time greater than time between meals, of course is a meal. Among the meals again it is lunch that seems to have the greatest effect on the human daily routine. While lower animals can be turned around completely and their biological clocks readjusted [Lobban, 1965], this is not easily done in humans. A contributing factor here are the social constraints that mankind has subjected itself to and that are difficult to alter. Lunch has become a social function during which ideas are discussed, business deals are talked over, friends are met, the workplace is left and in

many cases the family or home is visited. Especially in the last two aspects the lunch break or lunch time has become something to look forward to, to work up to, to condition one's mind towards. It cuts the working day in half, is something desirable and a great motivator. The administrator of an experiment that extends through the normal lunch period has to be aware of this conditon of the human.

Besides the psychological phenomenon of looking forward to that time of day, a physiological phenomenon exists. By about lunch time, breakfast, the earlier meal of the daily eating routine, has been digested and the body is running out of nutrients to keep working at the same pace as it has been during morning; the blood sugar level, for example, reaches a low level. [Haggard and Greenberg, 1935] The body is ready for a refill. It reminds the person of this fact by a feeling of hunger. Following this feeling, the person eats and experiences that the blood -eager to assimilate the nutrients from the food - fills the capillary system of the intestinal digestive tract rather than that of the brain. A decrease in the amount of blood in the brain area sets in and results in tiredness. Instead of following this feeling, however, and taking a siesta or a long noon break as is done in southern countries - the average person in industrialized countries continues to work following lunch. With reduced efficiency and reduced presence of mind and responses, the probability of errors, misjudgements,

erratic and useless movements increases, often resulting in accidents, injury or death; always resulting in reduced performance. As for the experimental set-up, the evaluation of various studies concerning effects of food intake on physical efficiency and mental acuity needs to be taken into acocunt. [Bjerner, Holm&Swensson, 1955; Haggard & Greenberg, 1935] In a long-term study (19 years), reading errors of meter readers (in Swedish gas works) were recorded (Figure 1); "reading errors were most frequent shortly after the midday break, during meal digestion." [Grandjean, 1970] Measurements of blood sugar level and respiratory quotient (Figure 2) during experimentally varied timings of meals indicated that low levels of these two indirect measures of efficiency can be avoided by increasing the number of meals to five and spacing them less far apart. [Haggard & Greenberg, 1935] [Figure 3] These findings have been confirmed by other experimetns.

Not for performance reasons initially, but for health reasons, dieticians have recommended the same timing of meals to relieve the body from the extremes of fullness and emptiness and have its metabolism instead perform at a rate that is more moderate and closer to constant.

The administering of in-between-meals during an experiment researching cyclical phenomena is therefore advisable since it will smoothe out physiologically induced variations of performance.

Any degradation of performance around the lunch time of the subjects - as observed in the data under discussion - should therefore not be attributed to an ultradian rhythm without keeping in mind that it is also the expected occurence following the circadian rhythm of the average subject.

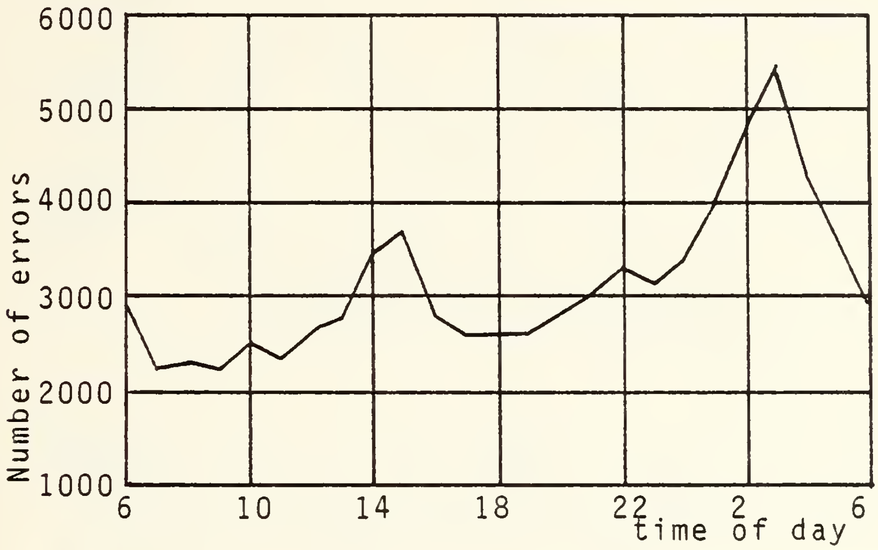

FIGURE 1. READING ERRORS OF GAS METER INSPECTORS, COMPILED OVER 19 YEARS. READING ERRORS REACH A MAXIMUM IN THE EARLY AFTERNOON (DURING THE DIGESTIVE PERIOD) AND DURING THE NIGHT WORK IN THE EARLY HOURS OF THE MORNING. [Bjerner, Holm, Swensson]

11. A further point of consideration in the design of an experiment during which measurements are taken repeatedly is the possibility of the subject's improvement due to a learning curve (LC). In the experiment under discussion the test problems were arranged in booklets of 48 pages each with 96 problems to a page. A subject, therefore, had to complete well

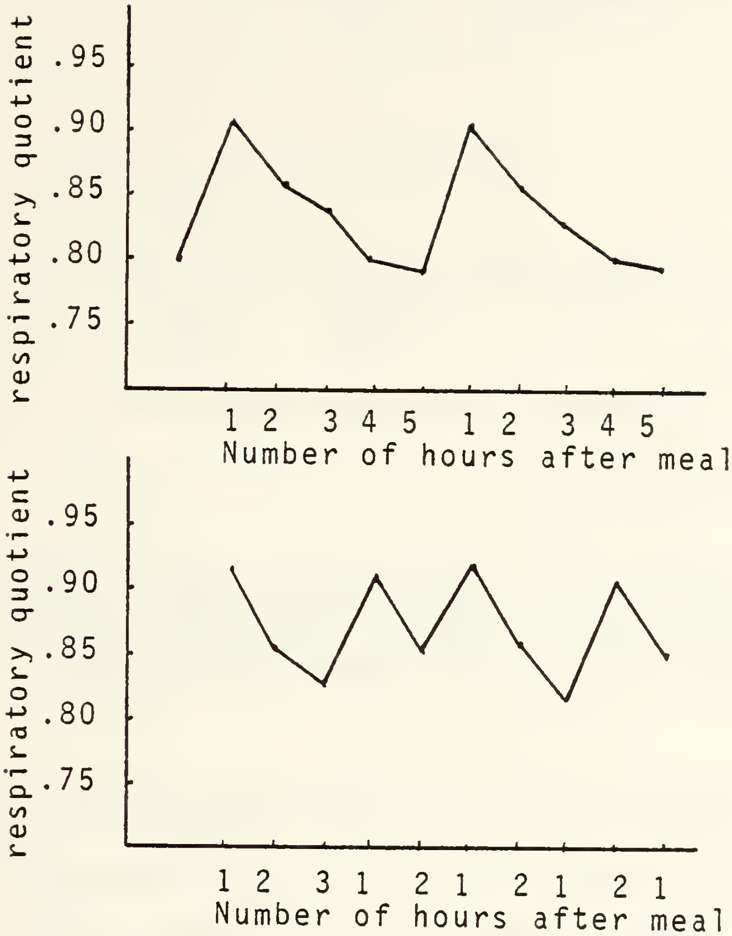

FIGURE 2. EFFECTS OF TWO MEALS (ABOVE) AND OF FIVE MEALS (BELOW). THE RESPIRATORY QUOTIENT CHANGES WITH THE BLOOD SUGAR AND CAN BE REGARDED AS AN INDIRECT MEASURE OF EFFICIENCY. WITH FIVE MEALS THE VALUES DO NOT DECREASE AS MUCH AS WITH TWO MEALS, SO THAT EFFICIENCY REMAINS HIGHER THROUGHOUT THE WORKING DAY. [Haggard & Greenberg]

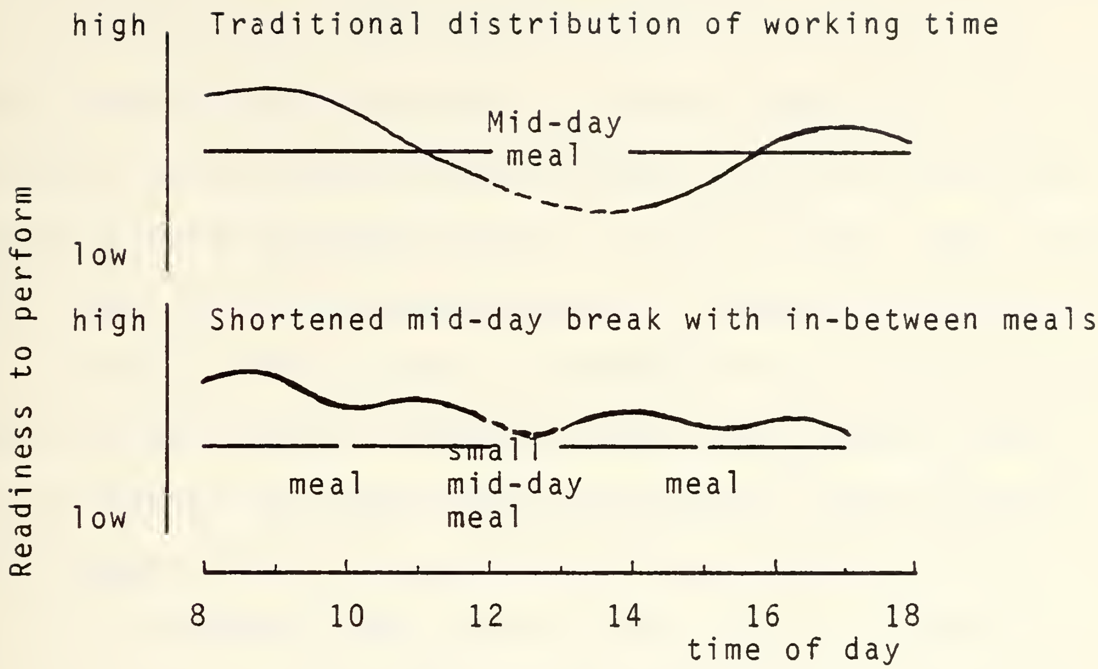

FIGURE 3. WORKING TIME, FOOD INTAKE AND READINESS TO PERFORM. DURING WORKING HOURS THE FOOD INTAKE SHOULD CONSIST OF TWO SNACKS AND A SMALL MID-DAY MEAL FOR REASONS OF HEALTH AND WORKING EFFICIENCY. [Grandjean]

over 4000 problems before she or he would return to the same problem. A LC due to same problems can consequently be ruled out. The existence of a LC in as far as finding a method to mark the answers in the best way was intended to also be ruled out - "subjects were given practice" [Klein] - the data, however, indicate a general performance increase (increase in number of problems solved) for the first few test intervals.

Last not least it must be understood that subjects will introduce their own bias into an experiment. Any effort which requires subjects to react in two ways that determine one outcome which is then measured: In Klein's experiment the Ss were asked "to make their same/different choices as quickly as possible without making errors." [Klein, 1979] This clearly burdens them with the responsibility to decide for themselves what is meant by "as quickly as possible" when weighed against "without making errors." The fact that they did make some errors indicates that they really made their same/different choices more quickly than possible without making errors. It needs to be examined which factor caused them to "accept" an error rate of 1.6. Had more emphasis been put on accuracy a zero error rate could have resulted, more stress on speed could have caused a higher error rate. This difference in emphasis can be looked for in the experimental instructions, in the way the instructions were given and possibly explained, but especially in the Ss' own understanding of the instructions, anticipation of a desired result and a personal weighting of the respective importance. This individually different trade-off of speed versus accuracy influences the outcome and needs to be considered.

II. STATISTICAL ANALYSIS

A. INTRODUCTION

The results of the analysis performed by the experimenters [Klein and Armitage] are as follows:

1. There is an ultradian rhythm of approximately 90 minutes in the performance data.
2. Verbal and spatial scores tend to have rhythms at this frequency which are 180° out of phase, i.e., verbal performance peaks when spatial performance is at its lowest level.

The statistical analysis given in this thesis and which is presented in the next sections comes to some distinctly different conclusions.

1. Most importantly the behaviour of the subjects is definitely inhomogeneous over the set of subjects. It will be shown that there are two sets of subjects which, within groups, show homogeneous behaviour in space (ensemble) and time, but between groups show disparate behaviour in frequency and time.
2. The only evidence for non-stationary behaviour in time in the two subgroups appears to be a four hour cycle, believed to be caused by a lunch-time effect.
3. There is very little evidence for a strong out-of-phase link between spatial and verbal behaviour.

The analysis is complicated by the fact that the usual assumptions of homogeneity in space and time are not valid, and that we are looking at eight bivariate (verbal-spatial) processes. Clearly there will be interactions between the inhomogeneities. The inhomogeneity between Ss will be presented first since it dominates the inhomogeneity in time.

It should be noted that the data set is not large enough to make very fine distinctions in the underlying behaviour mechanisms. Thus the conclusions beyond those of subject inhomogeneity are tentative.

B. EXPERIMENTER'S ANALYSIS

The experimenter converted the 32 scores of each S and each task to Z-scores through the process of normalization and performed a Fourier analysis. Power spectral peaks appeared at 4 hours, 96 minutes and at 37 minutes. Observed peaks were significantly larger than one would expect by chance (P .01) when tested through the use of two methods: In a "rank test" method the power spectra of each subject were replaced by their ranks and then ordered increasing from 1 to 16. A mean rank across eight Ss for each period was computed. Mean rank above 12.3 were found to be larger than expected by chance at the .01 level in a one-tailed test.

In a "randomization test" 500 random orderings of each S's scores for each task were subjected to Fourier analysis.

Using the mean root power of the random sequences for each subject, task and period the probability was determined of obtaining, by chance, a spectral value (root power) as large or larger than actually observed. Combining the probabilities across Ss it was determined for the group which periods had consistently larger peaks than would be expected by chance at the .01 level.

C. CRITIQUE OF EXPERIMENTER'S ANALYSIS

The scores are actually quite complex, since they are the number of correctly performed tasks in the three minute interval. No information is available on the actual number of tasks attempted. Thus the scores measure a mixture of the Ss's speed and error-proneness.

Standardizing by converting to Z-scores is also open to question. This standardization assumes normality. In fact the scores are quasi-count data, so that the mean and variance of the scores will not be independent. A square-root transformation, both to standardize the variance and induce normality would have been preferable. An alternative preferred in most of the analysis used in this thesis is to replace the data by the ranks within Ss. As will be seen in the two-way layout given below, this immediately shows up an inhomogeneity in the Ss's behaviour over time.

The spectral analysis is also subject to criticism. The technique is valid to test that the spectrum is not flat,

with the proviso that the maximum of the mean ranks must be used as the test statistic and the subjects must be homogeneous. Alternate tests for each spectrum are given below. However, rejection of flatness of the spectrum could indicate either that serial correlation is present in the data (very likely) or that there is a cycle, or both. The test results given below indicate very different results when the two subgroups of subjects are analysed separately.

D. PLOT OF RAW DATA (PERFORMANCE SCORES)

The analysis was begun by plotting the raw data. In this procedure, the spatial scores were marked along the horizontal axis, the verbal scores along the vertical axis. It was the goal of this plotting, which is shown in Figure 4, to display any correlation between the two elements of the paired test scores.

The Pearson product-moment correlation coefficient confirmed the visually apparent slightly negative correlation between the test scores of each of the eight subjects, as tabulated in Table I.

The approximate standard deviation for the coefficient in normal samples is $n^{-\frac{1}{2}}=32^{-\frac{1}{2}}=0.177$. Thus since we are looking at 8 of these values, the maximum value of -.48 is probably not significant. There is, however, a slight indication of negative correlation since all the correlations are negative. This question will be returned to later after

a spectral analysis is done on the 16 series (8 for verbal, 8 for spatial).

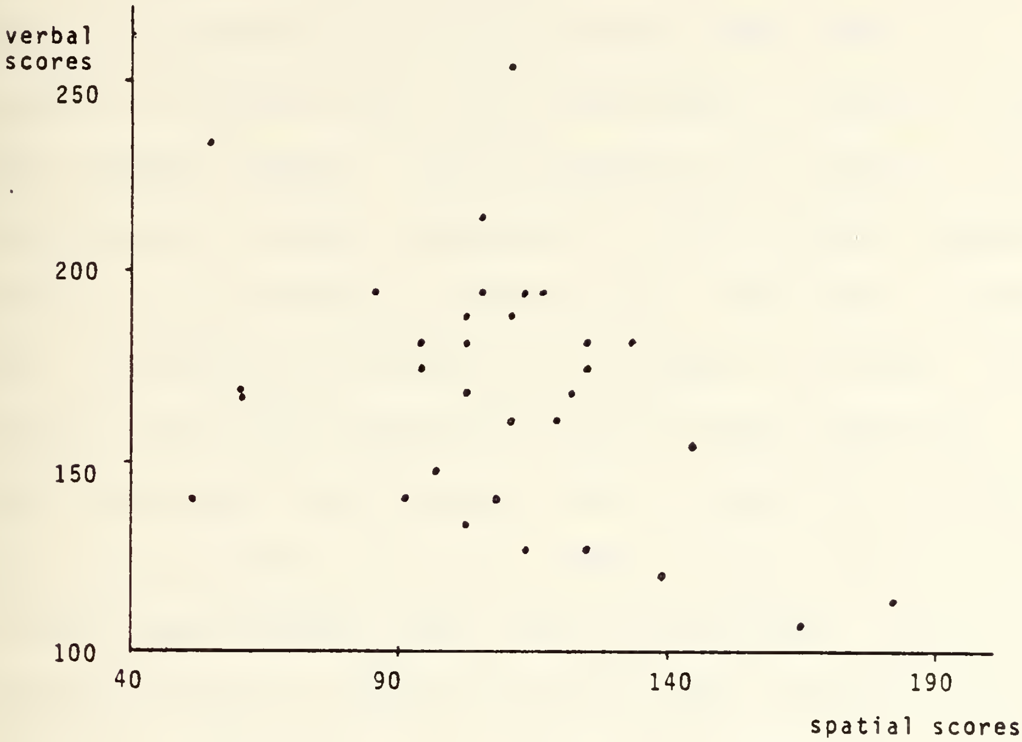

FIGURE 4. PLOT OF VERBAL SCORES VS. SPATIAL SCORES; REPRESENTATIVE PLOT; HERE: SUBJECT NO. 5. THE COMPUTED CORRELATION COEFFICIENT IS -.38 FROM TABLE I.

Subject	1	2	3	4	5	6	7	8
Correlation	-.34	-.48	-.36	-.10	-.38	-.28	-.12	-.11

TABLE I. CORRELATION COEFFICIENT OF VERBAL/SPATIAL PERFORMANCE SCORES.

E. FRIEDMAN'S TWO-WAY ANALYSIS OF VARIANCE

In the following section a nonparametric analog to the two-way analysis of variance procedure is performed. The data are presented in a kXn two-way table where k is the number of subjects and n is the number of trials (ranks). The subjects cannot - as the analysis will show - be considered a single, homogeneous random sample because of certain relationships between them. The rows in Tables II and III indicate subjects, the columns are ranks, i.e. the scores have been replaced by ranks within subjects. While the row totals are constant (sum of ranks) the column totals are only expected to be the same if there is no difference in the observations that lead to the ranks. Since the sum of deviations of observed column totals from expected column totals $k(n+1)/2$ is zero, the sum of squares of the deviations will reveal differences in observations, i.e. homogeneity or lack of homogeneity in the population of eight subjects. The first thing which is evident from this table is that the behaviour of the ranks classifies them into two groups. Thus note that at time period 7 subjects 1, 2, 3, 8 have low spatial ranks, while subjects 4, 5, 6, 7 have high ranks. Similar opposite behaviour is seen at other time periods, e.g. 12, 16, 19, 23, 30. The same effect can be seen from the verbal ranks (Table III); in fact, it shows up from the verbal ranks in most cases in the same time period as it does from the spatial ranks.

time period

Sbj	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	8	21	23	10	2	29	4	11	31	24	27	3	32	15	28	22	17	16	1	20	18	7	12	13	9	25	14	30	6	19	26	5
2	10	9	17	4	2	11	1	3	30	8	21	5	26	28	32	25	22	15	6	14	29	12	18	23	13	16	7	31	19	24	27	20
3	13	22	26	7	3	30	1	9	32	21	15	2	23	29	19	25	20	16	4	17	10	5	11	27	12	28	14	31	8	18	24	6
8	15	19	22	12	1	26	2	9	32	23	27	8	24	31	20	18	14	16	3	10	5	7	6	13	11	17	4	30	25	29	21	28
Rank Sum	46	71	88	33	8	96	8	32	124	76	90	18	105	103	99	90	73	63	14	61	62	31	47	76	45	86	43	122	58	90	98	59
4	24	22	25	32	28	23	30	9	16	26	17	19	14	27	13	5	8	6	31	10	20	11	7	21	29	18	12	2	3	1	15	4
5	12	6	17	30	29	13	31	16	18	26	20	21	14	22	11	3	5	15	32	19	25	23	9	10	28	7	8	2	1	4	24	27
6	9	8	1	29	28	12	30	11	19	27	20	21	25	22	15	7	17	10	32	14	23	24	5	26	31	13	16	2	3	4	18	6
7	5	7	20	28	29	22	32	10	23	18	24	26	13	8	14	3	6	9	31	19	11	17	4	15	30	27	25	16	21	1	12	2
Rank Sum	50	43	63	119	114	70	123	46	76	97	81	87	66	79	53	18	38	40	126	62	79	75	25	72	118	65	61	22	28	10	69	39
Total Rank Sum	96	114	151	152	122	166	131	78	200	173	171	105	171	182	152	108	111	103	140	123	141	106	72	148	163	151	104	144	86	100	167	98

TABLE II. TWO-WAY LAYOUT OF SUBJECTS AND THE RANKS OF SPATIAL TEST SCORES WITHIN EACH SUBJECT. WITHIN LIMITS THERE SHOULD BE SOME COHERENCE BETWEEN SUBJECTS. HOWEVER, ONE COULD HAVE CYCLIC PHENOMENA WITH (RELATIVELY) RANDOM PHASES AND THIS WOULD DECREASE COHERENCE OR CONCORDANCE BETWEEN SUBJECTS.

Sbj	time period 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	7	19	16	30	28	22	32	13	20	18	21	17	14	12	9	2	8	6	31	11	25	24	10	26	29	23	15	3	5	1	27	4
2	5	18	26	29	28	14	30	9	25	24	22	17	15	19	12	6	16	11	31	20	10	21	7	27	32	23	13	1	4	2	8	3
3	9	11	25	30	28	19	29	8	12	20	18	13	7	6	22	5	17	21	32	24	23	15	10	27	31	26	16	3	4	1	14	2
8	14	27	25	6	5	30	26	32	24	28	22	23	9	20	19	15	4	7	16	31	11	18	12	10	17	29	21	8	3	1	2	13
Rank Sum	35	75	92	95	89	85	117	62	81	90	83	70	45	57	62	28	45	45	110	86	69	78	39	90	109	101	65	15	16	5	51	22
4	4	11	16	20	5	29	27	19	32	21	30	2	26	25	28	24	22	23	1	15	17	7	9	14	12	18	6	31	10	13	8	3
5	6	8	13	11	3	20	1	9	32	19	26	5	27	28	24	16	29	30	2	25	22	12	10	15	21	18	23	31	7	14	17	4
6	7	3	18	17	4	27	1	9	32	28	22	2	13	14	19	6	30	29	5	21	10	8	26	20	23	24	12	31	15	25	16	11
7	9	7	10	11	2	21	1	6	30	24	26	4	16	18	19	22	14	27	5	29	23	13	15	25	20	31	28	32	12	8	17	3
Rank Sum	26	29	57	59	14	97	30	43	126	92	104	13	82	85	90	68	95	109	13	90	72	40	60	74	76	91	69	125	44	60	58	21
Total Rank Sum	61	126	149	154	103	182	147	105	207	182	187	83	127	142	152	96	140	154	123	176	141	118	99	164	185	192	134	140	60	65	109	43

TABLE III. TWO-WAY LAYOUT OF SUBJECTS AND THE RANKS OF VERBAL TEST SCORES WITHIN EACH SUBJECT.

This observation is crucial. We will now show that for overall eight subjects the Friedman statistic shows no common effect, but that this changes drastically if the statistic (or the equivalent Kendall's coefficient of concordance) is applied within subgroups.

The random variable S, the test statistic for the Friedman two-way analysis of variance, under the hypothesis of no difference between observations (ranks) therefore has a sampling distribution as follows:

$$S = \sum_{j=1}^{n} \left[R_j - \frac{k(n+1)}{2} \right]^2$$

Friedman's statistic F is a linear function of this random variable and is defined as:

$$F = \frac{12\ S}{kn(n+1)} = \frac{12}{kn(n+1)} \cdot \sum_{j=1}^{n} R_j^2 - 3k(n+1)$$

It can be treated as a chi-square variable with n-1 degrees of freedom as long as n is greater than 7. The results are given in Table IV. The rejection region for the test of the hypothesis of no difference in treatments (ranks) with n = 32 consists of all values greater than 45 or α = .05 and of all values greater than 52.2 for α = .01. This means - since all F-values of Table IV definitely fall into this region - rejection of the hypothesis of no difference at the .05 level.

At the .01 level the overall F-statistic for the spatial task falls just below the rejection region.

		overall	group 1238	group 4567
verbal	sum R_j^2	618448	167280	171418
	F stat	86.47	79.23	90.98
spatial	sum R_j^2	592805	172177	170624
	F stat	50.05	93.14	88.73

TABLE IV. LISTING OF THE FRIEDMAN STATISTIC FOR THE TEST OF THE HYPOTHESIS OF NO DIFFERENCE BETWEEN TREATMENTS (INDEPENDENCE OF SUBJECTS).

F. KENDALL'S COEFFICIENT OF CONCORDANCE

The above mentioned statistic F provides a possibilitiy to test the hypothesis of no difference between "treatments" - that results in ranks - at any level α. Actually of more importance is homogeneity or concordance between subjects.

A method to describe the concordance between subjects and to find a measure of relationship between rankings is the computation of Kendall's coefficient of concordance, W. This coefficient measures agreement between subjects, concordance between rankings or dependence of sample. It is analogous to the two-sample measurement of concordance; however, it does not range from -1 to 1, but only from 0 to 1. The reason for this is that among more than two samples

perfect discordance (as -1 indicates it in the two-sample case) cannot be defined. Zero means no agreement between subjects or independence of samples, 1 indicates perfect concordance. The statistic W is defined to be the ratio of S (as described in the previous section) to its maximum possible value k_{st} [Gibbons, 1971, pg. 251].

$$k_{st} = \frac{k^2 n(n^2-1)}{12} \qquad W = \frac{S}{k_{st}} = \frac{12\,S}{k^2 n(n^2-1)}$$

As can be seen from Table V the complete group of eight subjects is rather inhomogeneous, while the two previously identified subgroups 1, 2, 3, 8 and 4, 5, 6, 7 are much more homogeneous within groups. Table V also contains the necessary conversion of the statistic W to a statistic Y, a test statistic that allows a significance test at level α.

		overall	group 1238	group 4567
verbal	S	55072	27888	32026
	W	.31543	.63893	.73373
	Y	78.23	79.23	90.98
spatial	S	33917	32389	30968
	W	.19426	.74205	.70949
	Y	48.18	92.01	87.98

TABLE V. KENDALL'S COEFFICIENT OF CONCORDANCE AND THE TEST STATISTIC Y FOR SIGNIFICANCE TESTS.
$Y = k\cdot(n-1)\cdot W$ $\qquad$ $Y = k\cdot 31\cdot W$

This statistic is distributed chi-square with n-1 degrees of freedom. In a test of the hypothesis of independence between subjects large values of Y should fall in the rejection region. For n = 32 and α equal .05, values greater than 45, for α = .01 values greater than 52.2 have to be considered large. The result shown in Table V is therefore the same as that of Friedman's two-way layout discussed above.

G. FOURIER ANALYSIS OF RAW DATA (SASE VI)

The raw data without having been normalized or smoothed previously were subjected to Fourier analysis using the available SASE VI - program [LEWIS,1976].

The intent of this step of the analysis was to see if the power spectra (or amplitude) for the verbal and spatial performance scores as shown by Klein could be duplicated. Further it was desired to see if there were effectively any peaks in the spectra at any frequencies, and to compare the verbal and spatial spectra. The computation is as follows: let the set of data be indicated by X_i, i = 1, 2, . . ., n. The Finite Fourier transform of this data is

$$a(j) = \sum_{\ell=1}^{n} X_\ell \, e^{-i2\Pi(\ell-1)\cdot j/n}$$

$$; \quad j = 0, 1, \ldots, n-1$$

where $i = \sqrt{-1}$. The periodogram, an estimated amplitude spectrum of the data is defined to be

$$P(j) = \frac{|a(j)|^2}{2\pi n_{var}(X)}$$

$$; \quad j = 1, \ldots, \frac{n}{2} - 1$$

This estimates the spectral density of the sequence X_{O_i} given by

$$f+(\omega) = \frac{1}{\pi} \quad 1+2 \sum_{j=1}^{\infty} \rho_j \cos(j\omega) \quad ,$$

where the serial correlations ρ_j are defined to be

$$\rho_j = \mathrm{corr}(X_\ell X_{\ell+j}) \qquad j = 0, \pm 1, \pm 2, \ldots$$

We note that $f+(\omega) = \frac{1}{\pi}$ if the data are independent. More important are the sampling properties of the periodogram points $P_{(j)}$. These are approximately independent and exponentially distributed with mean value $f+(\omega)$. In particular if the spectrum is flat $f+(\omega) = \frac{1}{\pi}$ and the periodogram points are approximately i.i.d. exponential.

Before doing any formal tests for departure from a flat spectrum, it is worth looking at the spectra graphically. The spectra are smoothed by averaging over the eight subjects at each frequency f_j (or j), as was done by Klein. The numbers are given in Tables VI and VII, and the graphs in Figures 5 and 6. Klein's spectra are essentially reproduced

and the dominant feature in both is the peak at a frequency corresponding to a period of four hours. If this peak were significant, it would correspond to the lunch time effect noted in the introduction.

However, in the previous section we have seen that there is an inhomogeneity in the subjects. Therefore smoothed periodograms for the subjects 1, 2, 3, 8 and 4, 5, 6, 7 are shown for both verbal and spatial tests (Figures 5 and 6). The most interesting aspect of this split is that the strong similarity between the overall verbal and spatial periodograms disappears. In the spatial case (Figure 5) the dominant 4 hour peak is still apparent. However, the verbal periodogram for subjects 4, 5, 6, 7 does not have this peak. In fact, the verbal periodograms for the two groups are dissimilar, reinforcing the conclusion of heterogeneity in the sample of eight subjects.

Before any other inferences can be drawn from the periodograms, it is necessary to check the significance of the departure from a flat spectrum, since the peaks in the spectra may be due simply to sampling fluctuations. This is done in the next section.

	frequency j																	
Sbj	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	KS	WN2
1	0.205	0.567	0.009	0.004	0.389	0.076	0.111	0.205	1.102	0.120	0.027	0.081	1.024	0.400	0.607	0.011	1.008	1.250
2	0.487	0.996	0.014	0.101	0.693	0.167	0.193	0.018	0.203	0.452	0.297	0.160	0.454	0.016	0.673	0.023	0.860	0.743
3	0.006	0.358	0.093	0.294	0.441	0.253	0.095	0.247	1.295	0.082	0.188	0.116	0.950	0.158	0.211	0.295	0.741	0.707
8	0.376	1.512	0.081	0.044	0.220	0.261	0.203	0.234	0.430	0.127	0.366	0.282	0.371	0.083	0.215	0.262	--	--
Avg	0.266	0.858	0.049	0.111	0.436	0.189	0.151	0.176	0.758	0.195	0.220	0.160	0.700	0.164	0.427	0.148	0.870	0.900
4	0.553	1.125	0.625	0.272	0.477	0.355	0.037	0.071	0.086	0.001	0.411	0.092	0.442	0.056	0.098	0.468	1.376	3.098
5	0.435	1.239	0.170	0.029	1.011	0.323	0.157	0.211	0.328	0.024	0.285	0.015	0.366	0.059	0.027	0.510	1.237	2.982
6	0.659	0.623	0.029	0.048	1.054	0.363	0.029	0.007	0.110	0.011	0.852	0.082	0.676	0.011	0.074	0.613	0.908	1.137
7	0.116	1.156	0.376	0.605	0.535	0.104	0.094	0.092	0.194	0.225	0.211	0.096	0.545	0.017	0.032	1.074	1.303	2.330
Avg	0.441	1.036	0.300	0.239	0.769	0.286	0.079	0.095	0.180	0.065	0.440	0.071	0.507	0.036	0.058	0.666	1.206	2.387
Total Avg	0.355	0.947	0.175	0.175	0.603	0.238	0.115	0.136	0.469	0.130	0.330	0.116	0.604	0.100	0.242	0.407	1.062	1.750

TABLE VI. THE PERIODOGRAM VALUES FOR EACH SUBJECT AT EACH FREQUENCY $f_j=j/n$; $j=1, 2, \ldots, (n/2)-1$ FOR THE SPATIAL SCORES. THE AVERAGE (SMOOTHED) PERIODOGRAM IS GIVEN FOR EACH OF TWO SUBGROUPS AND FOR THE COMPLETE SET OF EIGHT SUBJECTS. THESE VALUES ARE GRAPHED IN FIGURE 5.

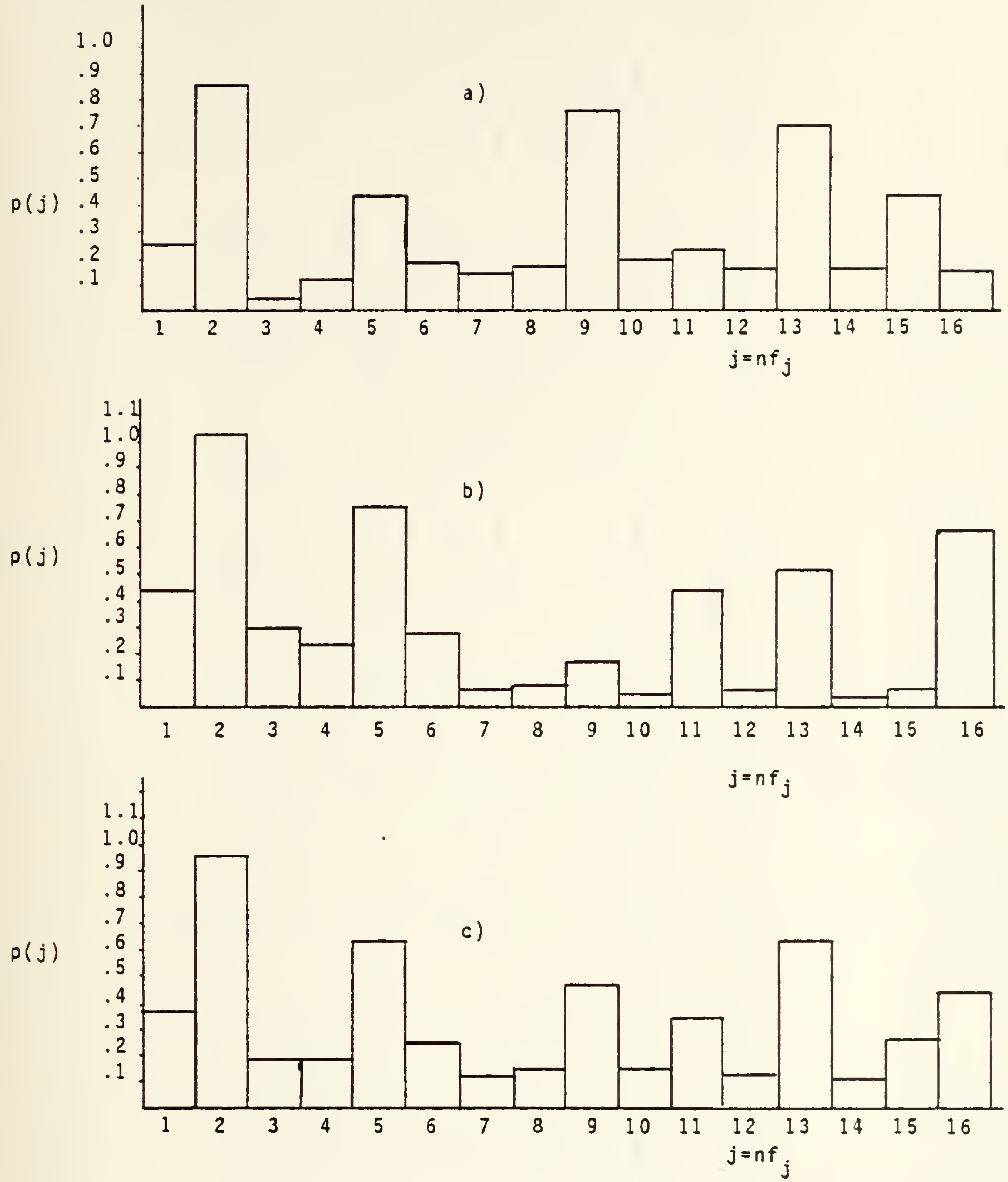

FIGURE 5. GRAPH OF SMOOTHED PERIODOGRAMS FOR SPATIAL SCORES OF THE TWO SUBGROUPS (a & b) AND FOR ALL SUBJECTS (c). IN ALL CASES THE 4 HOUR PERIOD COMPONENT (j=2) IS THE MAXIMUM PEAK, BUT IN THE CASE OF SUBJECTS 1238 THE 96-MINUTE PEAK IS NOT LARGE.

Sbj	frequency j 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	KS	WN2
1	0.112	1.726	0.136	0.111	0.578	0.339	0.012	0.060	0.148	0.009	0.541	0.054	0.623	0.001	0.024	0.919	1.270	2.588
2	0.390	1.479	0.389	0.291	0.487	0.375	0.031	0.079	0.077	0.041	0.406	0.047	0.490	0.012	0.006	0.670	1.439	4.055
3	0.169	1.810	0.326	0.353	0.514	0.302	0.021	0.101	0.144	0.002	0.262	0.112	0.240	0.082	0.019	0.955	1.593	4.806
8	0.481	1.168	0.009	0.303	1.465	0.041	0.149	0.123	0.037	0.056	0.419	0.017	0.099	0.121	0.044	0.806	1.760	4.374
Avg	0.288	1.546	0.215	0.265	0.761	0.264	0.053	0.091	0.102	0.027	0.407	0.058	0.363	0.054	0.023	0.838	1.516	3.956
4	0.866	0.490	0.328	0.028	0.563	0.011	0.052	0.054	0.444	0.052	0.062	0.242	0.747	0.224	0.770	0.002	0.786	1.212
5	0.469	0.500	0.125	0.169	0.360	0.155	0.178	0.207	0.815	0.280	0.104	0.389	0.655	0.255	0.260	0.028	0.494	0.293
6	0,165	0.445	0.298	0.197	0.578	0.108	0.352	0.134	0.540	0.021	0.362	0.564	0.768	0.076	0.316	0.022	0.516	0.236
7	0.950	0.605	0.049	0.473	0.293	0.297	0.083	0.081	0.346	0.075	0.137	0.113	0.942	0.310	0.083	0.196	0.936	1.172
Avg	0.613	0.510	0.200	0.217	0.449	0.143	0.166	0.119	0.536	0.107	0.166	0.327	0.778	0.216	0.357	0.062	0.683	0.728
Total Avg	0.450	1.028	0.208	0.241	0.605	0.204	0.110	0.105	0.319	0.067	0.287	0.192	0.571	0.135	0.190	0.450	1.099	2.342

TABLE VII. THE PERIODOGRAM VALUES FOR EACH SUBJECT AT EACH FREQUENCY $f_j=j/n$; $j=1, 2, \ldots, (n/2)-1$ FOR THE VERBAL SCORES. THE AVERAGE (SMOOTHED) PERIODOGRAM IS GIVEN FOR EACH OF TWO SUBGROUPS AND FOR THE COMPLETE SET OF EIGHT SUBJECTS. THESE VALUES ARE GRAPHED IN FIGURE 6.

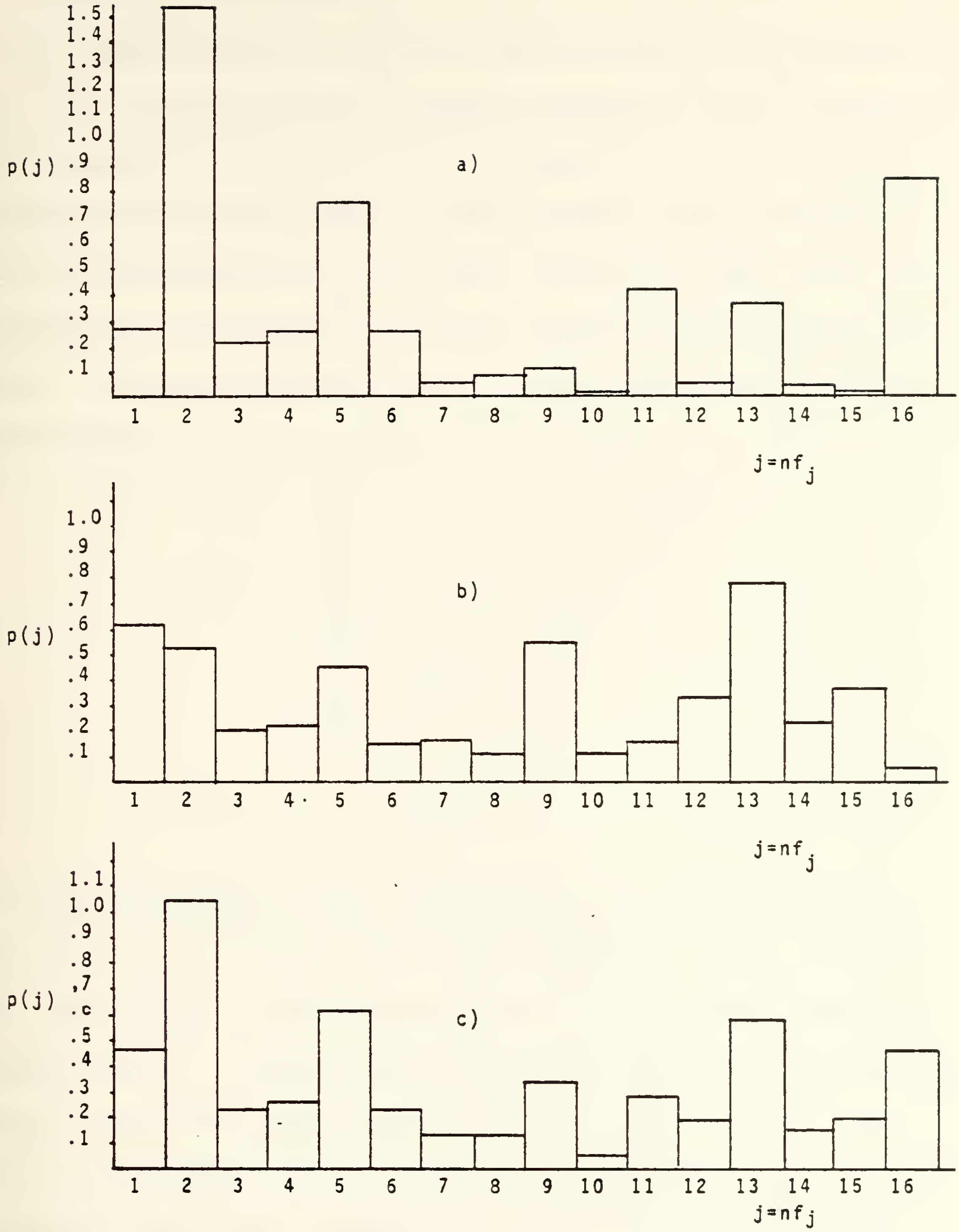

FIGURE 6. GRAPH OF SMOOTHED PERIODOGRAMS FOR VERBAL SCORES OF THE TWO SUBGROUPS (a & b) AND FOR ALL SUBJECTS (c). THE DIFFERENCE IN THE SPECTRA OF THE TWO SUBGROUPS IS STRIKING.

H. TEST FOR FLAT SPECTRUM

The pooled Fourier analysis performed by the experimenters [Klein, 1979] yielded a power spectrum in which dominant peaks occured at 4 hours, at 96 minutes and at 37 minutes. A Kolmogorov-Smirnov test for flat spectrum was applied to test the significance of all peaks together. This tests the independence of the 32 data points which go into making the spectra. The idea behind this is that since the p(j)'s are approximately i.i.d. and exponential, then

$$I(j) = \frac{\sum_{i=1}^{j} p(i)}{\sum_{i=1}^{\frac{n}{2}-1} p(j)} \quad ; \quad j = 1, \ldots, \frac{n}{2} - 2$$

are order statistics from a uniform (0, 1) distribution under the null hypothesis of a flat spectrum. [Cox & Lewis, 1966] But this is the canonical form of the usual tests of goodness-of-fit and thus the Kolmogorov-Smirnov and Anderson-Darling tests have been applied. [Cox & Lewis, 1966] The results are shown in Table VIII. For the verbal spectra it is striking that for subjects 1, 2, 3, 8 the K-S test is highly significant, while it is not significant for subjects 4, 5, 6, 7 at any level. Therefore the conclusion might be

that as far as verbal tests are concerned there is a strong lunch time effect in one case and not in the other. There is a much smaller chance that spatial tests were affected by the lunch time effect for the group 4, 5, 6, 7. This would confirm some opposition for verbal and spatial tests, with the qualification that it is a subject dependent result.

It would be difficult to extract from any of these spectra, jointly or spearately, any indication of a 90-min cycle in the data. The 90-min peak is big when the 4 hour peak is big, indicating that one sees a harmonic of the long cycle.

Subject	1	2	3	4	5	6	7	8
spatial KS	1.008	0.860	0.741	1.376	1.237	0.908	1.303	-----
two-sided test								
p=.80 KS=1.032	f	f	f	nf	nf	f	nf	
p=.90 KS=1.18	f	f	f	nf	nf	f	nf	
p=.95 KS=1.308	f	f	f	nf	f	f	nf	
p=.98 KS=1.464	f	f	f	f	f	f	f	
p=.99 KS=1.568	f	f	f	f	f	f	f	
verbal KS	1.270	1.439	1.593	0.786	0.494	0.516	0.936	1.760
two-sided test								
p=.80 KS=1.032	nf	nf	nf	f	f	f	f	nf
p=.90 KS=1.18	nf	nf	nf	f	f	f	f	nf
p=.95 KS=1.308	f	nf	nf	f	f	f	f	nf
p=.98 KS=1.464	f	f	nf	f	f	f	f	nf
p=.99 KS=1.568	f	f	nf	f	f	f	f	nf

f=flat nf=non-flat

TABLE VIII. KOLMOGROV-SMIRNOV TEST RESULTS FOR DIFFERENT p-LEVELS. FOR KS-VALUES HIGHER THAN THE ONES ASSOCIATED WITH THE p-LEVEL ON THE LEFT, THE HYPOTHESIS OF FLATNESS IS REJECTED. AT THE .05 LEVEL THIS IS ONLY THE CASE FOR SUBJECTS 3 AND 8 IN THE VERBAL SCORES.

I. KENDALL'S TAU

The previous sections have indicated some tendency of the performance in verbal and spatial tasks - as shown by the scores - to be negatively correlated and perhaps out of phase at some frequency. We will pursue this further here. Klein's hypothesis is that a cycle of about 96 minutes exists in the performance of a human and that verbal and spatial performance are about 180° out of phase, i.e. good verbal performance goes together with less good spatial performance and vice versa.

To test the second part of the hypothesis - the out-of-phase phenomenon - assuming for the time being that a cycle does exist, Kendall's tau coefficient was computed which is a measure of association between random variables from any bivariate population.

The reasoning behind using Kendall's tau as opposed to the Pearson product-moment correlation coefficient (as in paragraph D of this section) at this point is that tau is invariant under all transformations of X and Y for which the order of magnitude is preserved; this allows the use of ranks instead of raw data.

Perfect direct and indirect association between X and Y are reflected by perfect concordance and discordance, respectively. Concordance or agreement is a comparative type of association and means that large (small) values of X are associated with large (small) values of Y. [Gibbons,

1971, pg. 206] Discordance, the opposite, associates large X-values with small Y-values and vice versa.

The Null hypothesis in this test is that the X_i and Y_i are independent. The procedure to compute Kendall's tau, also called Theil's statistic, is as follows: Order the X, Y-pairs such that the X-values are in increasing order (this is shown in Table IX); the ranks of the X_i then run from 1 to n (1 to 32 in the experimental data under discussion). Now define P as the number of times the rank of a Y-value is greater than the rank of any of the following Y-values. Define Q as the number of times the rank of a Y-value is smaller than the rank of any of the following Y-values. P and Q then have to add to the total number of comparisons.

$$P = \#\text{ times } r(Y_i) > r(Y_j)$$

$$Q = \#\text{ times } r(Y_i) > r(Y_j)$$

$$j = i+1,\ i+2,\ .\ .\ .n$$

$$P+Q = \frac{n(n-1)}{2}$$

Now define K to be the difference P-Q. K is the difference of two random variables and is distribution-free. K is symmetric with a range of(-n(n-1))/2 to (n(n-1))/2 and an expected value E(K) = 0; the variance, sigma squared, equals (n(n-1)(2n+5))/18 under the Null hypothesis that X_i and Y_i

are independent. The reasoning for this is the permutation hypothesis which states that under assumption of independent identical distribution each of N! permutations of the ranks is equally likely. Define T to be K divided by (n(n-1))/2. For large n the random variable Z can be treated as a standard normal variable. It is a linear function of T and is defined as:

$$Z = \frac{3\sqrt{m(m-1)}}{\sqrt{2(2m+5)}} \cdot T$$

The results of the test of the hypothesis of independence of the X_i and Y_i is shown in Table X. Seven out of the eight subject results are not significant at the .05 level, not even at the .10 level. The hypothesis of independence is therefore accepted. This might be confounded by the frequency effect. Thus one could look at lagged correlations. This is pursued below in paragraph L of this section.

J. SPEARMAN'S COEFFICIENT R OF RANK CORRELATIONS

Another computation performed to be able to test the hypothesis of independence between the two different kinds of performance scores was finding Spearman's coefficient of rank correlation. The proceure is as follows: The X observations are replaced by their rank within the X_i-group; the Y observations are replace by their ranks in the Y_i-group. The data then consist of n sets of paired ranks and the

	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8
1	31	30	29	13	7	18	8	5
2	28	28	13	31	31	31	3	26
3	17	9	28	10	16	15	22	16
4	32	29	32	3	14	25	15	21
5	4	17	15	24	29	26	9	11
6	5	31	2	23	8	11	14	12
7	24	13	30	9	18	6	7	18
8	7	24	4	22	23	3	18	23
9	29	18	8	19	10	7	27	32
10	30	5	23	15	15	29	6	31
11	13	14	10	7	24	9	23	17
12	10	21	31	6	6	27	17	6
13	26	32	9	28	20	24	16	10
14	15	20	16	26	27	21	19	4
15	12	11	18	8	30	19	25	14
16	6	23	21	32	9	12	32	7
17	8	26	24	30	13	30	13	29
18	25	7	1	18	32	16	24	15
19	1	4	22	2	25	32	29	27
20	11	3	17	17	26	22	10	19
21	19	22	20	14	5	2	12	2
22	2	16	11	11	28	14	21	25
23	16	27	7	29	12	10	30	28
24	18	2	14	4	17	8	26	9
25	23	6	5	16	22	13	28	3
26	27	15	25	21	19	20	4	30
27	21	8	27	25	4	28	31	22
28	9	19	26	5	21	4	11	13
29	22	10	6	12	3	17	2	1
30	3	25	19	27	11	1	20	8
31	20	1	3	1	1	23	5	20
32	14	12	12	20	2	5	1	24

TABLE IX. RANKS OF VERBAL SCORES AFTER THE RANKS OF SPATIAL SCORES HAVE BEEN ARRANGED IN ASCENDING ORDER FROM 1 TO 32 (LEFT HAND COLUMN).

	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8
1	30	29	28	12	6	17	7	4
2	27	27	12	29	29	29	2	24
3	16	8	27	9	14	14	19	14
4	28	27	29	2	12	22	12	18
5	3	15	13	20	25	22	6	9
6	3	28	1	19	6	10	10	9
7	20	11	24	7	13	5	5	13
8	4	21	2	17	17	2	11	16
9	22	14	5	14	7	4	18	23
10	22	4	17	11	10	20	4	22
11	9	10	7	5	15	5	13	12
12	6	14	20	4	5	17	9	4
13	18	19	5	16	11	16	8	7
14	9	13	8	14	15	13	8	3
15	7	8	9	4	16	11	11	7
16	3	13	11	16	5	6	16	3
17	3	14	12	15	7	14	7	14
18	13	5	0	9	14	8	9	6
19	0	3	10	1	11	13	11	11
20	3	2	7	7	11	10	4	6
21	6	9	8	5	4	6	5	1
22	0	7	4	3	10	6	6	8
23	3	9	3	9	5	4	8	8
24	3	1	4	1	5	3	6	2
25	6	1	1	3	7	3	6	1
26	6	4	4	4	5	4	2	6
27	4	1	5	4	3	5	5	4
28	1	3	4	1	4	1	3	2
29	3	2	1	1	2	2	1	0
30	0	2	2	2	2	0	2	0
31	1	0	0	0	0	1	1	0
32	0	0	0	0	0	0	0	0
P:	279	324	283	264	296	288	235	257
Q:	217	172	213	232	200	208	261	239
K:	62	152	70	32	96	80	26	18
L:	1.01	2.46	1.14	0.52	1.56	1.30	0.42	0.29

TABLE X. LISTING OF P IN KENDALL'S TAU COMPUTATION AND STATISTIC L=K/SIGMA(K) FOR SIGNIFICANCE TEST. ONLY SUBJECT NO. 2 SHOWS SIGNIFICANT SIGNS OF DEPENDENCE; OVERALL INDEPENDENCE IS ACCEPTED.

coefficient R can be computed. Define R_i to equal rank(X_i) and define S_i to equal rank(Y_i); define D_i to equal the difference of R_i and S_i. The coefficient R is then defined to be:

$$R = 1 - \frac{6 \cdot \sum_{i=1}^{n} D_i^2}{m(n^2-1)}$$

"R measures the degree of correspondence between rankings, instead of between actual variate values, but it can still be considered a measure of association between the samples and an estimate of the association between X and Y in the continuous bivariate population." [Gibbons, 1971, pg. 226] As displayed in Table XI, with the exception of subject number 2, none of the subjects shows significant rank correlation to reject the hypothesis of independence between the verbal and spatial scores.

R1 :	-.19098	Z: -1.06	$Z=R(n-1)^{\frac{1}{2}}$
R2 :	-.42223	Z: -2.35	
R3 :	-.06745	Z: -0.38	mean of R :-.1504
R4 :	-.04857	Z: -0.27	std. dev. : .1498
R5 :	-.28556	Z: -1.59	
R6 :	-.18054	Z: -1.01	
R7 :	+.028776	Z: +0.16	
R8 :	-.03629	Z: -0.20	

TABLE XI. SPEARMAN'S COEFFICIENT OF RANK CORRELATION AND CONVERSION TO STANDARD NORMAL RANDOM VARIABLE.
WITH THE EXCEPTION OF SUBJECT NO. 2 NO SUBJECT SHOWS SIGNIFICANT RANK CORRELATION TO REJECT THE HYPOTHESIS OF INDEPENDENCE BETWEEN THE VERBAL AND SPATIAL PERFORMANCE SCORES.

K. RANK SUM MINUS EXPECTED RANK SUM UNDER H_o

Another way pursued to examine the hypothesized out-of-phase phenomenon was a plot of the sums of paired ranks minus the expected rank sum under H_o, $n+1 = 33$, vs. time period. Since under the hypothesis of Klein the performance of verbal and spatial tasks should be 180° out of phase. This would mean good performance (i.e. high ranks) in one task should go together with poor performance (i.e. low ranks) in the other. The expected value of this statistic is zero under Klein's hypothesis.

The procedure to develop this statistic is as follows: Replace the raw data by their ranks in each group of X and Y. Add the ranks of each pair X and Y and subtract 33, which is n+1 for the data under discussion. If the performance in the two different tasks really is out of phase the resulting values should be around zero. Very negative (or positive) results indicate an overall deviation to poor (or good) performance. Negative values can be expected in the beginning of the testing session due to a learning curve effect (LC) as well as later in the experiment due to fatigue.

Under Klein's composite hypothesis of a 96 minute cycle <u>and</u> an out-of-phase phenomenon for the performance on the two tasks one would expect this measurement to oscillate around zero following the 96 minute cycle. The result as displayed in Figure 7 shows, however, that - if anything at all - it follows a slow wave movement which is presumed to be induced by the lunch time effect.

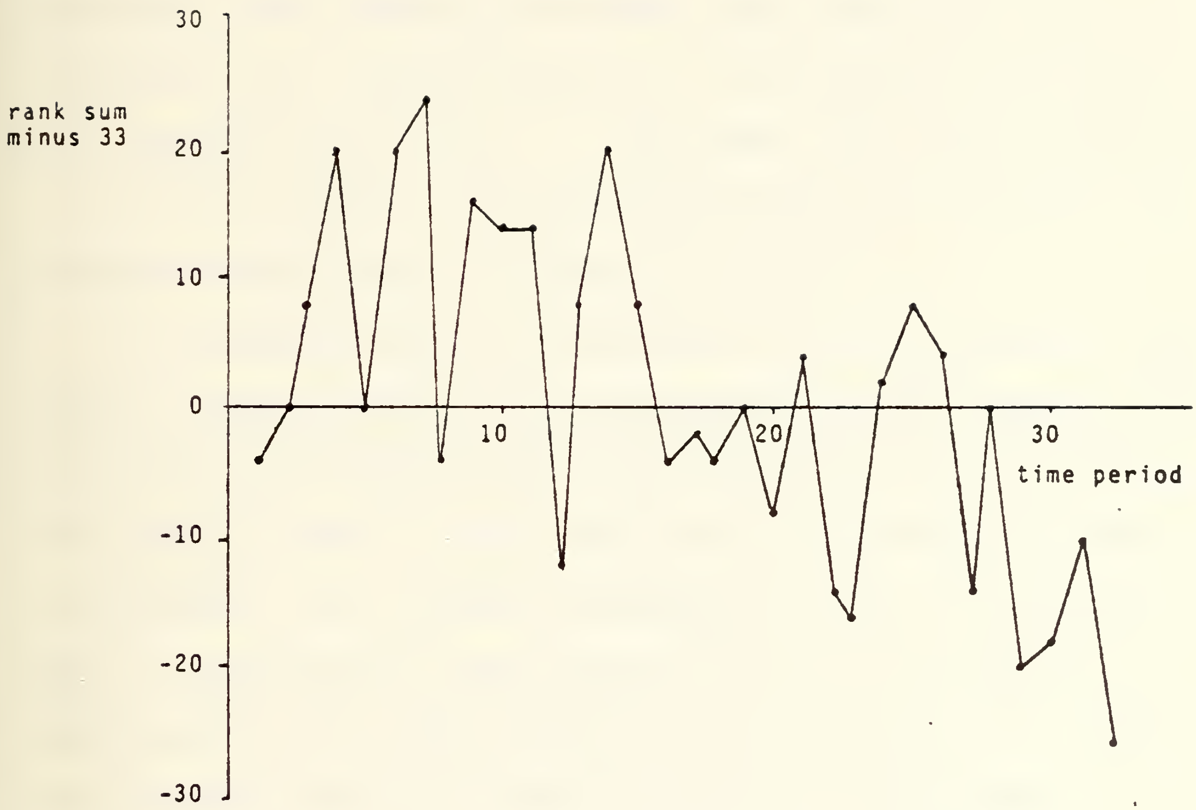

FIGURE 7. PLOT OF THE RANK SUMS OF BOTH VERBAL AND SPATIAL RANKS AT EACH INTERVAL MINUS THE EXPECTED RANK SUM UNDER KLEIN'S HYPOTHESIS OF OUT-OF-PHASENESS. LOW VALUES CAN BE EXPECTED AT BOTH ENDS OF THE GRAPH; DUE TO A LEARNING CURVE IN THE BEGINNING AND DUE TO FATIGUE AT THE END OF THE EXPERIMENT.

L. LAG SHIFTING OF RANKS

To investigate the out-of-phase hypothesis further, an analog to an auto-correlation procedure was devised. After replacing the raw scores by their ranks, one column of performance socre ranks was kept stationary and the other was shifted ("lagged") downwards for negative lag, upwards for positive; i.e. in this procedure spatial score (rank) position 1 was paired with verbal rank position 2 for positive lag 1; spatial position 1 was paired with verbal 3 for positive lag 2, etc... . For each lag in the range from 0 to ±15 the simple Pearson product-moment correlation was computed and plotted as shown in Figure 8.

The reasoning for this procedure is the following: Under Klein's hypothesis of out-of-phaseness one can expect a high positive correlation of two performance scores after about 3-4 shifts (lags). Since the scores were collected every 15 minutes, 3 to 4 shifts correspond to 45 to 60 minutes. With a hypothesized cycle period of 80 to 110 minutes, measurements that were initially out of phase would then be in phase and should be highly positively correlated.

The results for all subjects are similar to the representative plot of Figure 8, where a cycle of the hypothesized period cannot be found. The symmetry of the correlation plot, however, is striking: In all eight subjects the highest positive correlation is found after 8 to 10 shifts, indicating

a period of 4.5 to 5 hours. It could also - which is probably more realistic - simply indicate the positive correlation between poor scores at the beginning, after about 4 hours at lunch time and again 4 hours later due to fatigue.

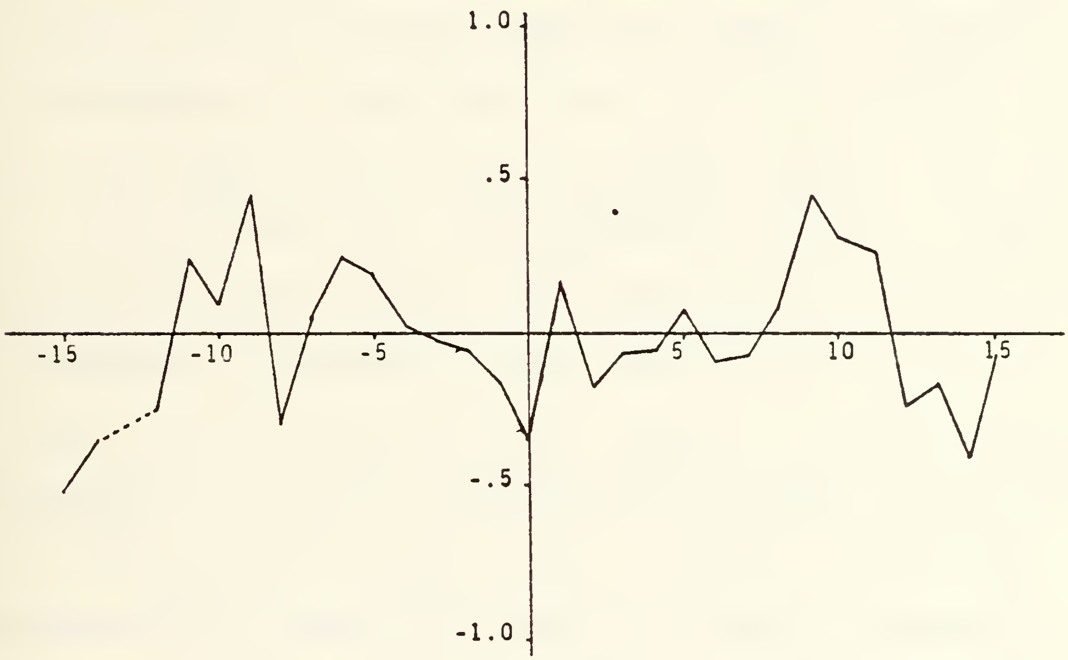

FIGURE 8. PLOT OF THE PEARSON PRODUCT-MOMENT CORRELATION FOR UP TO 15 SHIFTS IN POSITIVE AND NEGATIVE DIRECTIONS, RESPECTIVELY. THE EXPECTED STRONGLY POSITIVE CORRELATIONS AT ABOUT 3-4 SHIFTS CANNOT BE DETECTED. REPRESENTATIVE PLOT; HERE: SUBJECT NO. 3.

III. CONCLUSION

Considering the impact and importance of research in this part of cycle research as described in Section I, paragraph F, Klein and Armitage certainly investigated an interesting aspect of biological rhythmicity. Based on previous sleep research their hypothesis was a logical one. However, based on the research and analysis that went into the compositon of this thesis, their findings cannot be agreed to wholeheartedly. The inhomogeneity in the small group of their subjects seems to be a fact that has not been considered before. The influence of the 24-hour cycle, the circadian rhythm, on the test procedures and their timing seems sufficient to account for any rhythmicity detected by Klein and Armitage.

While further research into a possible ultradian rhythm is necessary and will be interesting, it needs to be evaluated if the impact of such a cycle - should it be found conclusively - is worth the effort in time and manpower. It also seems as if a useful predictability cannot be established since any possible influence of a possible ultradian rhythm is completely overshadowed by the variability among subjects.

It is suggested that a future experiment make use of some of the discussion objects of section I, paragraph I, that were not obvious to Klein and Armitage's experimental set-up.

LIST OF REFERENCES

Aschoff, J. (1963) Ann. Rev. Physiol. 25, 581

Bjerner, B.; Holm, A.; Swensson, A.; Brit, J. (1955) Industr. Med.

Bruce, V. G. (1965) In: Circadian Clocks (J. Aschoff, Ed.), pp. 125-138, North Holland Publishing Co., Amsterdam

Cohen, G. (1972) Percept. Psychophys. 11, 139

Cox, D. R.: Lewis, P. A. W. (1966) The Statistical Analysis of Series of Events, Wiley

Gibbons, J. D. (1971) Nonparametric Statistical Inference, McGraw-Hill

Grandjean, E. (1975) Fitting the Task to the Man, an Ergonomic Approach, pp. 80-86, 2nd ed., Taylor & Francis

Gross, M. (1972) Percept. Psychophys. 12, 357

Haggard, H. W.; Greenberg, L. A. (1935) Diet and Physical Efficiency, Yale University Press, New Haven

Harker, J. E. (1958) Biol. Rev. 33, I

Hartman, E. L. (1968) The 90-Minute Sleep-Dream Cycle, Archives of General Psychiatry, 18:280-286

Hartman, E. L. (1967) The Biology of Dreaming, Charles C. Thomas, Springfield, Ill

Jones, R. A. (1977) Self-fulfilling Prophecies, L. Erlbaum Associates

Klein, R.; Armitage, R. (1979) Rhythms in Human Performance, Science, Vol. 204

Kleitman, N. (1949) Physiol. Rev. 29, I

Kleitman, N. (1963) Sleep and Wakefulness, (Rev. ed.) University of Chicago Press

Lewis, P. A. W. (1976) SASE VI and The Statistical Analysis of Series in Events in Computer Systems, Naval Postgraduate School Technical Report NPS55LW76091

Lobban, M. C. (1965) Time, Light and Diurnal Rhythms, In: The Physiology of Human Survival, (Edholm and Bacharach, Editors), Acad. Press

Luce, Gay Gaer (1970) Biological Rhythms in Psychiatry and Medicine, U.S. Government Printing Office

Oatley, K.; Goodwin, B. C. (1971) The Explanation and Investigation of Biological Rhythms, In: Biological Rhythms and Human Performance, (W. P. Colquhoun, Editor), Acad. Press

Webster's New Collegiate Dictionary; G. & C. Merriam Co., 1976

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4.	Professor Douglas E. Neil, Code 55Ni Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
5.	Professor Peter A. W. Lewis, Code 55LW Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
6.	Professor William F. Moroney, Code 55Mp Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
7.	LCDR R. Carter, USN Navy Biodynamics Laboratory P. O. Box 29407 Michoud Station New Orleans, Louisiana 70189	1
8.	Dr. Alvah Bittner, PhD Navy Biodynamics Laboratory P. O. Box 29407 Michoud Station New Orleans, Louisiana 70189	1
9.	Torsten Andresen MARKTSTR. 6 2393 SOERUP West Germany	5